

LB Data Architecture - Schema & Models

- ELN Schemas
- ELN Spreadsheets design standards
- Lab Booster Data model
 - Context
 - Objective
- Entity-Relationship Diagram (ERD)
 - Entity-Relationship Diagram design
 - Entity dictionary
 - Entity-Relationship Diagram
- BigQuery
 - Batteries
 - Materials
 - Coatings
- Data mapping

This section aims to present the Data Architecture implemented for data flow in Lab Booster.

ELN Schemas

File Name	Data Model File
Agro	
Battery	
Coatings	
Seed Care	
Actizone	
HPC Flocculation	

ELN Spreadsheets design standards

Design is really important for user experience.

The user feedback about the first version of the ELN templates where more about the design than the content. They didn't really enjoy the "black and yellow" spreadsheets.

So for Coatings' Paint Formulation SS V2, we worked on defining design standards using Solvay's color palette as a base.

These standards are going to evolve according to the future needs.

Here are documentations about developed spreadsheet in PROD:

Agro		
Battery	Conductivity	
Battery	Mechanosynthesis	
Coatings	EP	
Seed Care	Formulation	
Seed Care	Results & Requests	

Lab Booster Data model

Overview

A **Data Model** represents the way data is structured in a dataset or a database, such as Lab Booster's data ocean.

The data model defines how the data lake or data ocean is connected to:

- The data input i.e. ELN, LIMS systems, connected instruments etc.
- The data output i.e. the WebApp DataLab in which users can access data

Context

As of mid-2023, each market in Lab Booster has its own data model i.e. its own way to structure data.

At each new project, connections to the data lake must be built again

Objective

Our aim is to have a common data model for all markets, to bring:

- Accelerated delivery of new projects
- Better performance
- Less maintenance

This page is divided into two sections

1. Entity-Relationship Diagram (ERD), which served as a basis to design the data model
2. Data model

Entity-Relationship Diagram (ERD)

Data Models are generally based on a diagram or schema called **Entity-Relationship Diagram** defining

- Entities i.e. a definable object or concept within a system
- Relationships i.e. how entities are related to one another

Building the ERD is a preliminary step to designing the actual data model to ensure that all required entities and relationships are accurately defined and represented.

This section is split into two parts

1. Entity-Relationship Diagram design
2. ERD mapping with R&I workflows

Entity-Relationship Diagram design

Entity dictionary

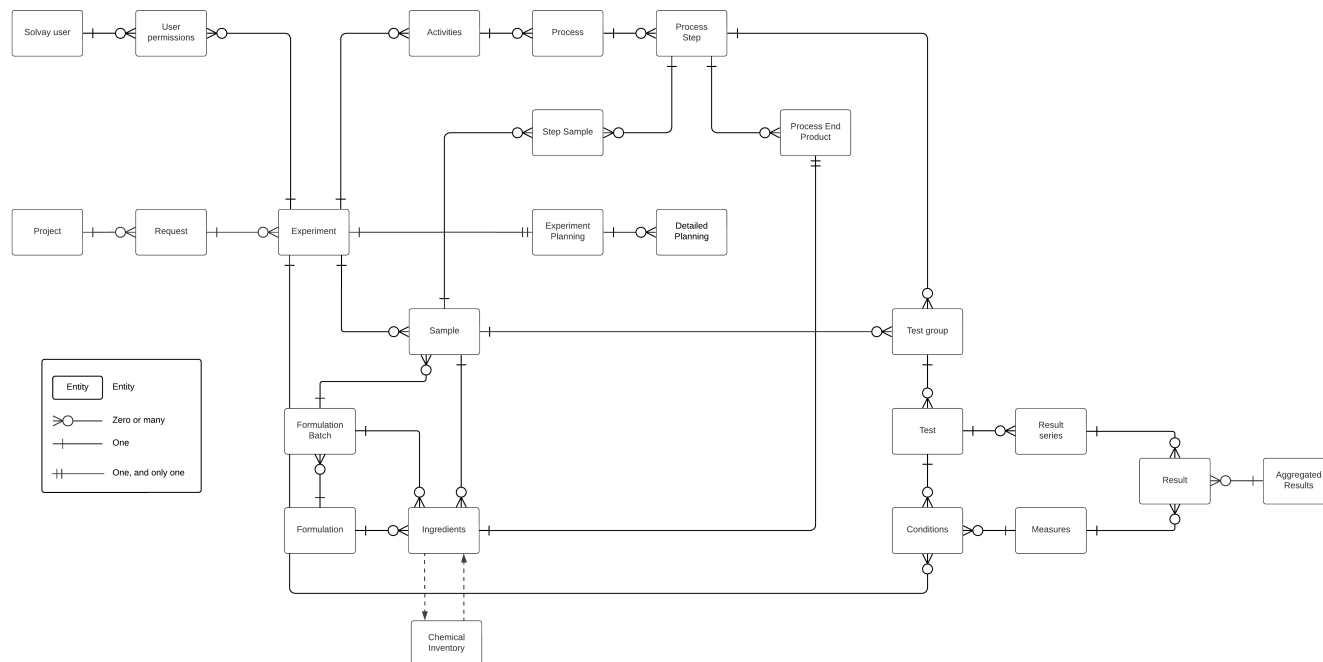
Entity	Definition	Example(s)
--------	------------	------------

Experiment	<p>A recording of a workflow performed in the lab by an operator at a given date to achieve an objective</p> <p>An Experiment includes</p> <ul style="list-style-type: none"> • <i>Activities</i> • <i>Samples</i> • <i>Tests</i> • <i>Request</i> • <i>Planning</i> 	<p>Experiments created and recorded in ELN IDBS, LIMS Labware LIMS Agilab...</p>
Solvay User	<p>A recording of the user that created the Experiment, including Solvay ID and email</p>	
User Permissions	<p>A setting determining what application options the user has access to</p>	
Request	<p>A recording of information provided by user requesting an <i>Experiment</i></p> <p>A Request includes</p> <ul style="list-style-type: none"> • Request date • <i>Sample</i> information • Information on user making requests 	<p>Requests for BioMatTech - Biodegradability testing include</p> <ul style="list-style-type: none"> • Request name • Requestor name • Request date • Priority • Status • Test method required • Sample name • Sample ID • Sample status • etc.
Planning	<p>A recording of when the <i>Experiment</i> is supposed to be performed</p> <p>A Planning includes</p> <ul style="list-style-type: none"> • <i>Tests</i> or <i>Activities</i> expected date • <i>Results</i> availability date 	<p>Planning in Novocare - Méréville Request & Results includes</p> <ul style="list-style-type: none"> • Expected application (of slurries & powders on seed) date • Operator performing the application
Activity	<p>A group of <i>Processes</i> performed in the lab in a specific order</p>	<p>In Novocare - Méréville Request & Results, two Activities are found, Application and Testing</p>
Process	<p>A group of <i>Process Steps</i> performed in the lab in a specific order</p>	<p>In BatMat -Mecanosynthesis, the Mecanosynthesis Process is defined by several successive <i>Process Steps</i></p> <ol style="list-style-type: none"> 1. Jar Preparation 2. Milling 3. Drying 4. Calcination 5. Finishing
Process Step	<p>A recording of tasks performed in the lab, defined by its name and date</p> <p>A Process Step includes</p> <ul style="list-style-type: none"> • <i>Conditions</i> in which it is carried out • Input and output <i>Step Samples</i> • <i>Tests</i> performed during Process Step • <i>Process End Product</i> <p>A Process Step follows a Standard Operating Procedure (SOP)</p>	<p>In Aroma - Fermentation the Growth Process Step is defined by the date on which it is performed and includes</p> <ul style="list-style-type: none"> • <i>Conditions</i> - Scale, Temperature, pH... • Input <i>Step Samples</i> - Starter media and Substrate • Output <i>Step Samples</i> - Sample #, Date and Time • <i>Tests</i> - Optical Density and Glucose analysis • <i>Process End Product</i> - Growth media
Process End Product	<p>The chemical output of a <i>Process</i>, defined by its name and date</p> <p>Process End Product characteristics include composition, aspect, mass and/or volume...</p> <p>A Process End Product can be registered as a new <i>Ingredient</i> for other <i>Formulation (Batch)</i> or <i>Process Steps</i></p>	<p>In Aroma - Fermentation, the Process End Product of the <i>Process Step</i> "Bioconversion" is vanillin</p> <p>In Novocare - Méréville Formulation Recipe, the Process End Product of the <i>Formulation Process Step</i> is a formulation</p> <p>In BatMat - Mecanosynthesis Jar Slurries, Amorphous Precursors and Raw Calcined Products are Process End Products</p>

Ingredient	<p>A chemical product, defined by its name and unique ID and recorded in an inventory</p> <p>Ingredient characteristics include date, batch number, supplier, physical state (liquid/solid), density, color...</p> <p>An Ingredient can be:</p> <ul style="list-style-type: none"> • A <i>Formulation Batch</i> • A <i>Sample</i> • A <i>Process End-Product</i> 	<p>In Aroma - Fermentation, the substrate Ferulic acid is an Ingredient</p> <p>In Novocare - Méréville Request & Results, Slurries and Powders are Ingredients</p> <p>In BatMat - Mecanosynthesis Jar Precursors, Slurries, Amorphous Precursors and Raw Calcined Products are Ingredients</p>
Formulation	<p>A combination of chemical products defined by the <i>Ingredients</i>, the <i>Ingredients</i> target proportions and its name</p> <p>Formulation characteristics include total number of chemical products, target concentration, target volume, calculated density...</p>	<p>In Novocare - Méréville Request & Results, a Recipe is a Formulation and is defined by name, ID and label.</p> <p>Characteristics include Number of products, Products, Recipe unit, Recipe Price, Calculated Recipe Density...</p>
Formulation Batch	<p>A combination of chemical products defined by the <i>Ingredients</i> actual proportions, its name, unique ID and date</p> <p>Formulation Batch characteristics include total number of chemical products, actual concentration, total volume, density, container (vessel, jar, bottle)...</p> <p>A Formulation Batch is a <i>Formulation</i> that has been created in the lab</p>	<p>In Novocare - Méréville Request & Results, a Batch of Recipe is a Formulation Batch and is defined by name, ID and label</p> <p>Characteristics include Recipe selection, Actual Weight (of Products)</p>
Sample	<p>A part of a substance or component that is taken from the whole substance or component, defined by its name, unique ID and date</p> <p>A Sample can come from</p> <ul style="list-style-type: none"> • An <i>Ingredient</i> • A <i>Formulation Batch</i> • A <i>Process End-Product</i> • A <i>Request</i> <p>A Sample can be used for</p> <ul style="list-style-type: none"> • A <i>Test</i> • A <i>Process Step</i> <p>See <i>Step Sample</i> for Samples taken during a <i>Process Step</i></p>	<p>Samples come from</p> <ul style="list-style-type: none"> • An <i>Ingredient</i> : Inoculum in BioMatTech - Biodegradability • A <i>Formulation Batch</i> : Batch of Recipe in Novocare - Méréville Formulation • A <i>Process End-Product</i> : Finished Product in BatMat - Mecanosynthesis <p>Samples are used for</p> <ul style="list-style-type: none"> • A <i>Test</i>: Batch of Recipe to characterize at t0 in Novocare - Méréville Formulation • A <i>Process Step</i>: Growth mass used in Bioconversion <i>Process Step</i> in Aroma - Fermentation
Step Sample	<p>A part of a substance or component that is taken from the whole substance or component in relation to a <i>Process Step</i>, defined by its name and date</p> <p>A Step Sample can be</p> <ul style="list-style-type: none"> • An input for the <i>Process Step</i> • An output of the <i>Process Step</i> 	<p>In Aroma - Fermentation, Step Samples are taken throughout the three <i>Process Steps</i> to monitor the chemical reactions</p>
Sample Test Plan	<p>A <i>Planning</i> defined for a set of <i>Samples</i>, defined by its name and the timing</p> <p>The Sample Test Plan characteristics include total number of <i>Samples</i>, <i>Tests</i> to perform ...</p> <p>A Sample Test Plan can apply in the context of</p> <ul style="list-style-type: none"> • A <i>Process Step</i> • A <i>Request</i> • A <i>Planning</i> 	<p>In Novocare - Méréville Formulation the Sample Test Plan defines when Samples should be taken during an ageing <i>Process Step</i></p> <p>It is defined by</p> <ul style="list-style-type: none"> • Protocol name • Initial storage date • Number of <i>Samples</i>
Test Group	<p>A group of <i>Tests</i> performed on the same <i>Sample</i></p>	<p>Characterization tests (OD manual, OD density and Glucose) performed during the Growth <i>Process Step</i> in Aroma - Fermentation for a Test Group</p>

Test	A measure of <i>Sample</i> behavior when a procedure is carried out	Tests performed in BatMat - Mecanosynthesis include Particle size test, SEM test, Lumisizer test, H NMR test, P31 NMR test, Li7 NMR test, Discrete value test
Measure	A property that can be measured A Measure can serve both a <i>Condition</i> and/or a <i>Result</i>	pH is a <i>Condition</i> in Aroma - Fermentation and a <i>Result</i> in BioMatTech - Biodegradability
Conditions	A variable or setting defined by the operator for <ul style="list-style-type: none"> • A <i>Test</i> and affecting its <i>Result</i> • A <i>Process Step</i> 	In BioMatTech - Biodegradability, Conditions for the Dry matter <i>Test</i> include Empty aluminium cup weight In Aroma - Fermentation, Conditions of the Growth Process Step include Scale, Temperature, pH...
Results	The outcome of a <i>Test</i> performed on a <i>Sample</i> in specified <i>Conditions</i> Results can take the form of <ul style="list-style-type: none"> • A numerical value • A set of numerical values (i.e. curve) • A non numerical value (i.e. observations) 	A pH value is a Result of a biodegradability <i>Test</i> in BioMatTech - Biodegradability A conductivity curve is a Result of a conductivity <i>Test</i> in BatMat - Conductivity Observations are a Result of a Look after Attrition <i>Test</i> in Novicare - Méréville Request & Results
Results Series	A set of <i>Results</i> , obtained at different time intervals, for a <i>Test</i> performed in the same <i>Conditions</i> on the same <i>Sample</i>	
Aggregated Result	A <i>Result</i> obtained by aggregating <i>Results</i> from several <i>Tests</i>	In Aroma - Fermentation, the maximum amount of vanilin produced during the Bioconversion <i>Process Step</i> is an Aggregated Result as it aggregates several vanilin concentration measure <i>Results</i> In Novicare - Méréville Request & Results, averages calculated from two different <i>Test Results</i> are Aggregated Results

Entity-Relationship Diagram



ERD mapping with R&I workflows (WIP)

Three types of R&I workflows were identified

- Formulation workflows
- Synthesis workflows

- Analysis workflows

This was done in order to ensure that the ERD defined accomodates all types of R&I workflows.

The mapping done for different workflows is summarized in the table below.

GBU/F- R&I	Workflow name	Workflow type	Mapping status	Link to mapping	Documentation - Data capture
Novecare GBU	Seed Care Formulation	Formulation	Done	Seed Care mapping	ELN template
Novecare GBU	Seed Care Request & Results	Formulation	Done	Seed Care mapping	ELN template
Battery Platform	Mecanosynthesis	Synthesis	Done	Mecanosynthesis mapping	ELN template
Aroma Performance GBU	Fermentation	Synthesis	Done	Fermentation mapping	ELN spreadsheet mockup
BioMatTech Platform	Biodegradability	Analysis	Done	Biodegradability mapping	LIMS spreadsheet mockup
Specialty Polymers GBU	Aging, Mechanical, Thermal	Analysis	Ongoing		
Specialty Polymers GBU		Synthesis	To do		
Novecare GBU	Agro	Formulation	To do		
Novecare GBU	EP Coatings	Synthesis	To do		
Novecare GBU	Paint Coatings	Formulation	To do		
Corporate R&I	Solvent platform - Solubilization		To do		
Corporate R&I		Analysis	To do		
Green Hydrogen Platform	Conductivity	Analysis	To do		

BigQuery

New Data Model of ALB Data Mart (Exposition layer): https://app.genmymodel.com/api/projects/_k07o4lBOEe29ie0vpi-P5A/diagrams/_k07o4oBOEe29ie0vpi-P5A/svg

Data Mapping to Data Mart:

The following BigQuery datasets are all staging as per the data convention explained previously.

For more ETL (extraction, transformation, loading) details, please refers to: [App Lab Booster \(ALB\) - Data](#)

Batteries

File Name	Data Mapping File
Agro	
Coatings	Version 2:
	Version 1:
Battery	
Materials	
Seed Care	